PAPER

Continuous Music-Emotion Recognition Based on Electroencephalogram

Nattapong THAMMASAN^{†a)}, *Nonmember*, Koichi MORIYAMA^{†*}, *Member*, Ken-ichi FUKUI[†], *Nonmember*, and Masayuki NUMAO[†], *Member*

SUMMARY Research on emotion recognition using electroencephalogram (EEG) of subjects listening to music has become more active in the past decade. However, previous works did not consider emotional oscillations within a single musical piece. In this research, we propose a continuous music-emotion recognition approach based on brainwave signals. While considering the subject-dependent and changing-over-time characteristics of emotion, our experiment included self-reporting and continuous emotion annotation in the arousal-valence space. Fractal dimension (FD) and power spectral density (PSD) approaches were adopted to extract informative features from raw EEG signals and then we applied emotion classification algorithms to discriminate binary classes of emotion. According to our experimental results, FD slightly outperformed PSD approach both in arousal and valence classification, and FD was found to have the higher correlation with emotion reports than PSD. In addition, continuous emotion recognition during music listening based on EEG was found to be an effective method for tracking emotional reporting oscillations and provides an opportunity to better understand human emotional processes.

key words: music, emotion, electroencephalogram

1. Introduction

Emotion is a crucial factor in human-computer interaction. An emotion reflects a mental state and psychophysiological expression. Scientists realize that an emotion has strong connection with physiological signals, including brainwaves, and emotion-brain research has become a highly active research area. An electroencephalogram (EEG) allows the feasible and cost-effective investigation of emotion by analyzing electrical activities along the scalp with high temporal resolution. Based on the neural correlate studies of emotion using EEG data, various algorithms, e.g. fractal dimension (FD), power spectral density (PSD) and discrete wavelet transform, have been proposed to extract meaningful information from EEG data and construct models to recognize human emotion [1], [2].

Although emotion can be evoked by various types of stimuli, including pictures, videos, or even voluntarily, music is one of the most frequently used materials in emotion research because of several benefits. For example, music is considered as an extraordinary material to elicit emotion powerfully and evoke a wide variety of emotions [3]. Music

also enables a study of time courses of emotional processes. Furthermore, using music in EEG-based emotion recognition also has a promise of potential applications such as music therapy [4], implicit multimedia tagging [5] and retrieval [6]. Thus, we focus on emotion while listening to music.

Emotion while listening to music can change over time, especially for long-duration music. Studies have found that peripheral physiological reactions of listeners during music perception change over time relative to emotional states that were elicited continuously by music [7]. In an fMRI study [8], brain activation differences between the first 30 seconds and the remaining 30 seconds of musical excerpts were found. The following EEG study using the same stimuli also confirmed the differences of cortical activities [9]. Considering the dynamic process of music emotion, a considerable amount of works in music emotion variation detection have been done in the last decade by utilizing musical features [10]. However, information from musical pieces does not always reflect the listener's emotional states.

A huge number of works have been proposed to estimate emotional states during music listening by using EEG and peripheral signals. However, the previous studies did not consider the emotion variation because the chosen musical excerpts were relatively short (less than one minute). Most existing music emotion recognition using EEG have been based on a single emotion annotation for one musical excerpt [1]. Indeed, duration of music in the real world is generally longer than one minute.

In this study, we propose a method to extract emotion-related time-varying information from EEG signals during music listening and create an emotion recognition model based on the time-varying "ground truth". This study focuses on continuous emotion annotations in a single song rather than employing an entire song-level annotation. Because duration of songs is sufficiently long to change the emotion, and the capability of our approach to capture emotional oscillation in songs is also investigated. Importantly, emotion when listening to music is subjective, *i.e.*, the same piece of music can induce different emotions in different listeners. Therefore, rather than relying on the predefined emotion labels indicated by another listener, we gather self-annotated emotion labels from the actual listeners.

The remainder of this paper is organized as follows. In Sect. 2, we briefly review the study related to emotion recognition based on physiological reactions. We describe our

Manuscript received June 29, 2015.

Manuscript revised December 3, 2015.

Manuscript publicized January 22, 2016.

[†]The authors are with the Institute of Scientific and Industrial Research, Osaka University, Ibaraki-shi, 567-0047 Japan.

^{*}Presently, with the Graduate School of Engineering, Nagoya Institute of Technology, Japan.

a) E-mail: nattapong@ai.sanken.osaka-u.ac.jp

DOI: 10.1587/transinf.2015EDP7251

methodology and experimental setting in Sect. 3. In Sect. 4, we present experimental results, and we discuss the results in Sect. 5. Finally, in Sect. 6, we conclude this paper.

2. Related Work

2.1 Dimensional Emotion Space

Emotional representation models have been proposed to describe emotion systematically. The dimensional approach defines models based on the principle that human emotions could be represented as any points lying in two or three continuous dimensions. One of the most prominent models is the arousal-valence emotion model proposed by Russell [11]. In this bipolar model (Fig. 1), valence is represented by the horizontal axis indicating the positivity or negativity of emotions. The vertical axis represents arousal which describes the activation levels of emotions. In this study, we have employed the arousal-valence emotion model to represent human emotions because it has been shown to be an effective and reliable model for recognizing emotion during music listening [12].

2.2 EEG Bandwave

In healthy adults, changing from one cognitive state to another leads to the alteration of the amplitudes and frequencies of EEG signals. The electrical activities of the brain are classified according to rhythms and defined in the terms of delta(δ), theta(θ), alpha(α), beta(β), and gamma(γ) from low to high frequencies [13]. It is occasionally referred to as EEG bandwaves because each wave lies within a specific range. The frequency ranges of these brain waves and their association with normal human activities are summarized in Table 1.

2.3 EEG-Based Emotion Recognition

Studies on neural correlates of emotion have found evidence

of emotion-influenced changes in EEG signals. Evidence of higher activity in the left frontal lobe of the brain in comparison with the right hemisphere while subjects were experiencing positive emotions, prominently in α band power, have been reported [14], [15]. Moreover, Sammler *et al.* [9] found increases in the frontal midline theta power while subjects were listening to pleasant musical excerpts. Based on these discoveries, computational and machine learning algorithms have been applied to EEG signals to achieve high-performance in emotion estimation and prediction [1].

Various types of materials have been used to elicit emotion. Images from the International Affective Picture System (IAPS) [16] were utilized in research on the identification of emotion from EEG data [17]. In addition, videos were also employed to evoke targeted emotions [18], [19] and self-elicitation has also been performed [20]. Different algorithms have been introduced to extract informative features from EEG signals, such as PSD [20], higher order spectra [21], higher order crossings [22], and discrete wavelet transform [23].

Music has also been used as stimuli in EEG-based emotion recognition. As music-emotion recognition based on EEG and peripheral signals is still in its infancy, researchers are aiming to identify two or more finite classes of arousal and valence although emotion can be represented continuously in dimensional space [1]. Bos [24] used sound clips from the International Affective Digitized Sounds (IADS) [25] and images from IAPS to classify emotional states into positive-arousal, positive-calm, negative-arousal, and negative-calm states. Bos achieved an accuracy of 92.3%. Khosrowabadi *et al.* [26] introduced kernel density estimation and Gaussian mixture model probability estimation to extract features from EEG data and classified six categorical emotions using a Bayesian network, multi-layer perceptron (MLP), one-rule, random tree and a radial basis function. They accomplished inter-subject accuracies of 90%. Lin *et al.* [27] used pre-labeled music, and joy, sadness, anger, and pleasure emotions could be discriminated at a performance rate of 85%. Sourina *et al.* [28] utilized self-emotion reporting after listening to specific sound clips and

Fig. 1 The arousal-valence emotion space (redrawn from Russell [11]) with axis labels

Table 1 Comparison of EEG bandwave summarized from [13]

Band	Frequency range (Hz)	Association
Delta	0.5–4	Deep sleep
Theta	4–8	Consciousness slips
		Drowsiness
		Unconscious material
Alpha	8–13	Creative inspiration
		Deep meditation
		Relaxed awareness
Beta	14–30	Eye closing
		Active thinking, Attention
		Motor behavior
Gamma	30 +	Focusing on the outside world
		Solving concrete problems
		Sensory processing
		Certain cognitive motor function

emotion labels in selected clips from IADS, and positive-high-aroused, positive-low-aroused, negative-high-aroused, and negative-low-aroused emotions were discriminated at an accuracy of 84.9% for arousal classification and 90.0% for valence classification.

Although many methods to estimate the emotional states that utilize EEG data have been proposed, these studies have primarily used pre-emotion-labeled music pieces obtained from standard libraries, where such emotions are reported by labeling by experts or another person. Realizing the fact that emotion during music listening is subjective and can change over time, we applied a technique to combine temporal continuous annotation and self-reporting.

3. Research Methodology

3.1 Participants and Materials

Fifteen males between 22 and 30 years of age (mean = 25.52, SD = 2.14) participated in the experiments. All subjects were mentally healthy students of Osaka University. None had formal music education.

Our music collection is a set of MIDI files comprised of 40 instrumental pop songs having different instrument and tempo. By using MIDI files, any additional emotions contributed by lyrics can be eliminated. MIDI files also enable musical feature investigation and application to music composition which are considered as our future works.

3.2 Data Collection

Our data collection software was developed using Java. The experiment began with a questionnaire regarding personal information and musical preferences. Then, the subject selected 16 MIDI musical excerpts from the 40-song MIDI collection using the software. We asked the subject to choose eight familiar songs and eight unfamiliar songs for further investigation of music familiarity in our future work. Our software provided a function to play short samples of songs to facilitate recognizing familiarity, where 1-3 referred to low familiarity (unfamiliar songs) and 4-6 denoted high familiarity (familiar songs).

We placed a Waveguard EEG cap[†] on the subject's head in accordance with the 10–20 international system to measure electrical activities along the brain. We selected 12 electrodes, *i.e.*, Fp1, Fp2, F3, F4, F7, F8, Fz, C3, C4, T3, T4, and Pz, out of 21 available electrodes (Fig. 2) as these electrodes are located close to the frontal lobe which plays a crucial role in emotion regulation [4], [14]. The sampling frequency was set to 250 Hz. The impedance of each electrode was less than 20 k Ω . A notch filter, a type of bandstop filter that reduces a narrow range of frequencies, was applied to reduce the 60-Hz electrical power line artifact. Brain signals were transmitted to a Polymate AP1532^{††} amplifier and

Fig. 2 Position of selected electrodes in accordance with 10–20 international system

then visualized by its software, APMonitor^{†††}.

Later, the selected music clips were presented as sounds synthesized by the Java Sound API's MIDI package^{††††}. The average duration of each song was approximately two minutes. Each song trial ended with a 16-second silent rest to reduce the effects from the previous song. The subjects were instructed to close their eyes and minimize body movement while wearing the EEG cap and listening to music. After completing all listening sessions, the subject removed the EEG cap and proceeded to an annotation session. During the session, the subject listened to the same songs and annotated his/her emotions perceived in the previous session continuously by clicking on corresponding points in the arousal-valence emotion space shown on a monitor. Arousal and valence were recorded independently as numeric values from -1 to 1 . A brief guideline of arousal-valence emotion space which included Fig. 1 was given throughout annotation session to acquaint the subjects with the arousal-valence model.

3.3 Data Preprocessing

To filter out unrelated artifacts, a bandpass filter was applied to extract only 0.5–60-Hz EEG signals. We utilized EEGLAB [29] to identify and reject distinct artifact-contaminating data automatically. The rejection of epochs of continuous EEG data was implemented using a function in EEGLAB. In addition, eye-blinking related artifacts were removed from EEG signals by applying the independent component analysis (ICA) signal processing method [30]. ICA decomposes multivariate signals into independent non-Gaussian subcomponents. Lacking a dedicated electrooculogram, we adopted an artifact removal technique into using the components of Fp1 and Fp2 electrodes instead because the two frontal electrodes were positioned nearest to the eyes and are obviously influenced by eye-blinking artifacts. Fi-

[†]<http://www.ant-neuro.com/products/waveguard>

^{††}<http://www.teac.co.jp/industry/me/ap1132/>

^{†††}Software developed for Polymate AP1532 by TEAC Corporation.

^{††††}<http://docs.oracle.com/javase/7/docs/technotes/guides/sound/>

nally, we associated EEG signals with the emotion labels annotated by the subjects via timestamps.

3.4 Feature Extraction Algorithms

EEG signals were processed to retrieve informative features using two approaches, FD and PSD. The calculations were performed by MATLAB analysis tools. We applied a sliding window segmentation technique to analyze temporal data and track emotional fluctuation. The window size was defined as 1000 samples, which was equivalent to 4 seconds. In this study, the overlap between one sliding window and the consecutive window was set to zero.

3.4.1 Fractal Dimension

FD values characterize the complexity of time-varying signals. Higher FD values of EEG signals reflect the higher activity of the brain [31]. FD values are typically employed in affective computing research, including emotion recognition based on EEG [28], because of their simplicity and informative characteristics that properly indicate brain states. In this study, we applied the Higuchi algorithm [32] to calculate time series FD values directly in the time domain.

Given a time series $X(i)$ where $i = 1, \dots, N$, a new series $X_m^k(i)$ can be constructed by the following definition:

$$X_m^k : X(m), X(m+k), \dots, X\left(m + \left\lfloor \frac{N-m}{k} \right\rfloor k\right), \quad (1)$$

where k is the interval time and $m = 1, 2, \dots, k$ is the initial time. For example, assuming that the series has $N = 100$ elements and $k = 3$, then the series is separated into three series as follows:

$$\begin{aligned} X_1^3 &: X(1), X(4), X(7), \dots, X(97), X(100) \\ X_2^3 &: X(2), X(5), X(8), \dots, X(98) \\ X_3^3 &: X(3), X(6), X(9), \dots, X(99). \end{aligned} \quad (2)$$

Then, the length of the series X_m^k is defined as:

$$L_m(k) = \frac{1}{k} \left[\frac{\sum_{i=1}^{\lfloor \frac{N-m}{k} \rfloor} |X(m+ik) - X(m+(i-1)k)|}{\frac{\lfloor \frac{N-m}{k} \rfloor k}{N-1}} \right], \quad (3)$$

where the term $\frac{\lfloor \frac{N-m}{k} \rfloor k}{N-1}$ represents a normalization factor. The length of time interval k , denoted $\langle L(k) \rangle$, is obtained by averaging all the sub-series lengths $L_m(k)$. The following relationship exists:

$$\langle L(k) \rangle \propto k^{-\text{FD}}. \quad (4)$$

The FD is a gradient of a logarithmic plot between k and its associated $\langle L(k) \rangle$ and the calculated slope.

3.4.2 Power Spectral Density

Over the last few decades, the PSD analysis of EEG data has

been a typical approach to investigate the relevance of affective states and brainwaves [1]. PSD indicates signal power in specific frequency ranges. This method is based on fast Fourier transform, which is an algorithm to compute the discrete Fourier transform and its inverse. This transformation converts data in the time domain to the frequency domain and vice versa. It is widely used for numerous applications in engineering, science, and mathematics.

In this research, each EEG signal is decomposed into five frequency ranges (delta, theta, alpha, beta, and gamma) using the PSD approach. The PSD values were calculated using MATLAB Signal Processing Toolbox[†]. As PSD represents signals in the continuous frequency domain, we needed to calculate a feature that represents the overall characteristics of a specific frequency range. As the feature, we used the average power over the given frequency band calculated using the *avgpower* function in the toolbox.

3.5 Emotion Classification

In this research, emotional arousal was classified as high or low. Similarly, emotional valence was classified as positive or negative. Because of the application of sliding window technique, subjects' annotated emotion labels could vary within one window. We unified emotional tagging to the window by the *majority* method, *i.e.*, emotional tagging in one window was set as high or positive if the number of positive instances was greater than that of the negative instances (and vice versa).

For the classification of emotion, we trained two models to identify arousal and valence classes independently by employing three types of classification algorithms: support vector machine (SVM), MLP, and C4.5. The SVM and MLP are typical algorithms in brain-computer interaction research. C4.5 is superior for speed of learning. All classification algorithms were implemented using the Waikato Environment for Knowledge Analysis (WEKA) library [33].

4. Experiments and Results

4.1 Experimental Setup

After retrieving experimental data, we applied feature extraction algorithms to the data from each electrode. As a result, we obtained 12 features by FD value calculation. In contrast, the PSD approach produces 60 features.

Previous reports have indicated that asymmetries of features from symmetric electrode pairs can be used as informative features to classify emotions [19], [27], [28], [34]. One plausible reason is that the asymmetry indexes might suppress underlying artifact sources that contributed equally to hemispheric electrode pairs [35]. Therefore, we added asymmetry indexes to our original features as additional features. These additional features were the differences of feature values from the left hemisphere electrodes and the symmetric electrodes in the right hemisphere, *e.g.*, the difference

[†]<http://www.mathworks.com/help/signal/ref/dspdata.psd.html>

Fig. 3 Average arousal and valence classification accuracy for all subjects; error bars denote standard deviation and stars indicate significant difference compared to chance levels

between the power of Fp1 and that of Fp2 in PSD features. There were five symmetric electrode pairs; consequently, we obtained 17 FD-value features and 85 PSD features.

Then, we trained a subject-dependent recognition model and tested it using data from one subject. We adopted the 10-fold cross-validation method to evaluate each subject's model to obtain overall classification results. Note that the non-overlapping characteristic of the adjacent sliding window in our experiment avoided classification bias caused by similar training and testing instances.

Traditional methodologies have primarily neglected emotional changes over time. In other words, the informative features were extracted at the song-length level. Those traditional models were trained on aggregated instances from multiple songs. To simulate such conventional methods, we adapted our methodology by expanding the size of the sliding window to the full length of the song. We trained the model with the song-length data using the same feature extraction and classification algorithms. The overall label of the window is produced by a majority of annotations.

4.2 Chance Level

As our research relies on subject annotations, an unbalanced data set could be obtained. In other words, if a subject labels his/her perceived emotion primarily as a positive valence, the number of positive valence instances would be higher than the number of negative instances. This asymmetry would lead to misinterpretation of classification results. Therefore, we introduce a new indicator, *chance level*, as a benchmark to evaluate models. The chance level, or random guessing level, is defined by the majority class of the training data. For example, in a training set consisting of

60% positive samples and 40% negative samples, the chance level would be 60%.

4.3 Results of Emotion Classification

Emotion classification accuracy in each 10-fold cross-validation was defined as the proportion of correctly classified test instances (true positives and true negatives) among the total number of instances in the test set. The average emotion classification accuracy for all subjects is shown in Fig. 3. All approaches that considered the dynamics of emotions outperformed the chance level for arousal recognition significantly ($p < 0.01$). Classification by FD value features with the SVM achieved the best relative result (82.8%, $SD = 8.1\%$). In this case, the chance level was 62.0% ($SD = 6.6\%$). It should be noted that data from a subject who annotated only single class of arousal was removed from arousal classification to avoid any bias.

Similarly, valence classification performance was superior to the chance level. Again, the FD approach was superior to the chance level significantly ($p < 0.01$) regardless of classification algorithm (SVM achieved the highest accuracy at 87.2%, $SD = 5.9\%$), while PSD gave better results compared to the chance level only with SVM and MLP classifiers. In valence classification, the chance level was 72.9% ($SD = 13.0\%$).

Compared with traditional approaches, using statistical paired t-test, all of our methodologies demonstrated superior performance. In particular, for all algorithms, considering emotional oscillation improved the performance of arousal classification significantly ($p < 0.01$). Valence recognition by any feature extraction or classification technique also achieved higher results ($p < 0.05$).

Fig. 4 Arousal and valence annotations from subject No. 4 and their estimation by the model constructed with SVM and FD values from all instances (horizontal axis represents the order of songs selected by the subject; data are plotted in time order)

5. Discussion

This research focuses on continuous emotion recognition relying on continuous self-reported emotion labels. The improved performance of continuous emotion recognition over the traditional approach of using song-level labels are promising but leave room for discussion. Empirical results also showed that each feature extraction algorithm and classifier achieved different results, which we analyze and discuss in this section. To investigate, we studied the association of features with the reported emotional states. Furthermore, we examine whether our approach could track emotional variation by visualizing estimated emotion and then comparing it with subject-reported emotion.

According to the obtained results, our proposed methodology that considers emotion variation and applies a sliding window technique outperformed traditional methods. It is possible that the results of conventional approaches could suffer because of the limited training examples. To compensate for this, multiple sessions are required to elicit different types of emotions. Multiple sessions incur a time cost because of a large number of resting periods between sessions. On the other hand, the proposed method utilizes fewer songs to construct an emotion recognition model, which is a more practical technique in real-world applications. Continuous annotation enables temporal data segmentation and provides larger amounts of data to analyze. According to our results, temporal data partitioning has enhanced the efficiency of emotion recognition empirically.

To investigate the correlates of extracted features with emotion reports, we computed Pearson product-moment correlation coefficients between each EEG feature and the

Table 2 Top-5 features with the highest absolute averaged correlations (over all subjects) with the valence and arousal ratings

		FD value	PSD
Arousal	Fp2	0.1405	Fp2- γ 0.0705
	F7	0.0858	F7- β – F8- β 0.0665
	Fp1 – Fp2	-0.0776	Fp2- δ 0.0581
	C3	0.0727	F3- θ – F4- θ -0.0572
	F4	0.0702	C3- δ -0.0562
Valence	F4	0.1040	F8- α 0.0804
	F3 – F4	-0.0966	T4- α 0.0777
	Fz	0.0536	C3- α 0.0755
	Pz	-0.0462	Fp2- α 0.0735
	F8	0.0357	T3- α 0.0683

numerically reported arousal and valence for each subject separately. The resulting correlations were then averaged over all subjects to produce overall correlations. For FD value and PSD features, 5 features with highest absolute averaged correlations with arousal and valence are summarized in Table 2. The involvement of these features was partly consistent with previous literature. The FD asymmetry at F3–F4 was comparable with the finding that the frontal FD asymmetry at AF3–F4 can recognize valence with high accuracies [36]. In a work using DEAP dataset [19], FD values of F4 and F7 were among the 4 channels selected by the Fisher discriminant ratio channel selection method to classify emotions using FD and HOC features [31]. Further, the relevance of beta asymmetry at F7–F8 and theta asymmetry at F3–F4 to arousal was consistent with their involvement in top-5 ranked asymmetric features to classify emotions in previous work [27], [34].

The classification results suggest that FD value features outperform PSD. The generally higher arousal correlations of FD value features compared to PSD features may have

contributed to the superior performance of models using FD value features for arousal recognition. Similarly, valence recognition with FD value features could achieve better results compared to PSD features because of their slightly higher absolute correlations. This evidence coincides with results from previous studies in the field of EEG-based affective computing [37], [38], which reported that the FD approach is superior to PSD in recognizing affective states because of the superior ability to analyze the non-linear behavior of the brain. Note that the SVM achieved better results than the other classifiers, *i.e.*, C4.5 and MLP, and that similar results were also obtained in previous works [1].

Examining whether the model could track variation of self-reported labels is also informative. To illustrate this, we used data from subject No. 4 because of the obvious fluctuations in the annotation. We trained the recognition model with all instances and compared the estimation of emotion from the trained model with the annotated labels. The models were trained by using FD and SVM because of their success in our experiment. The results for arousal and valence recognition are shown in Fig. 4. The horizontal axis shows the order of songs selected by the subject from songs 1 to 16. Emotional reporting oscillations are observable for some songs. According to the results, the emotion recognition model could handle the distinct shifts of emotion during some songs. For example, the model changes the estimation of arousal from low to high for songs 2 and 12. Furthermore, the proposed model could track the trajectory of valence shift from negative to positive while listening to song 1 and could track conversely during song 5. However, the results shown in Fig. 4 need to be interpreted carefully. The model was trained on all the available instances in the dataset; hence, it reflects the maximum in the capability of the model to capture emotions, which could be slightly higher than the results from cross-validation. Moreover, the model relies heavily on subjective annotation whose capability of reflecting *real* emotions is limited by the subject's own emotional self-awareness.

In addition, data available to investigate tracking of emotional reporting oscillations were relatively limited; *i.e.*, the annotation oscillations were found in eight of sixteen songs on average over all subjects. Therefore, further experimentation with select songs with high emotional oscillation may be necessary to confirm the capabilities of our approach. This is considered as our future work.

6. Conclusion

In this work, we have presented a study of continuous music-emotion recognition using EEG based on the hypothesis that emotions evoked when listening to music are subjective and vary over time. Experiments were performed by focusing on self-reporting and continuous emotion annotation in the arousal-valence space. The results showed that our approach outperformed traditional approaches that did not consider emotional changes over time for arousal and valence recognition, especially when classifying emotions

with FD value features and SVM classifier. We also found that the emotional ground truth had higher correlation with FD value than PSD. Finally, the models constructed through our approach were found to display a satisfactory capacity for tracking reporting oscillations of subjects' emotions while listening to music.

References

- [1] M.-K. Kim, M. Kim, E. Oh, and S.-P. Kim, "A review on the computational methods for emotional state estimation from the human EEG," *Comp. Math. Methods in Medicine*, vol.2013, pp.1–13, 2013.
- [2] R. Jenke, A. Peer, and M. Buss, "Feature extraction and selection for emotion recognition from EEG," *IEEE Trans. Affective Computing*, vol.5, no.3, pp.327–339, 2014.
- [3] S. Koelsch, *Brain and Music*, Wiley-Blackwell, 2012.
- [4] S. Koelsch, "Brain correlates of music-evoked emotions," *Nat. Rev. Neurosci.*, vol.15, no.3, pp.170–180, 2014.
- [5] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE Trans. Affective Computing*, vol.3, no.1, pp.42–55, 2012.
- [6] J. Eaton, D. Williams, and E. Miranda, "AFFECTIVE JUKEBOX: A confirmatory study of EEG emotional correlates in response to musical stimuli," *Proc. ICMC/SMC 2014*, pp.580–585, 2014.
- [7] O. Grewe, F. Nagel, R. Kopiez, and E. Altenmüller, "Emotions Over Time: Synchronicity and Development of Subjective, Physiological, and Facial Affective Reactions to Music," *Emotion*, vol.7, no.4, pp.774–788, 2007.
- [8] S. Koelsch, T. Fritz, D. Cramon, K. Müller, and A.D. Friederici, "Investigating emotion with music: An fMRI study," *Human Brain Mapping*, vol.27, no.3, pp.239–250, 2006.
- [9] D. Sammler, M. Grigutsch, T. Fritz, and S. Koelsch, "Music and emotion: Electrophysiological correlates of the processing of pleasant and unpleasant music," *Psychophysiology*, vol.44, no.2, pp.293–304, 2007.
- [10] Y.H. Yang and H.H. Chen, "Machine recognition of music emotion: A review," *ACM Trans. Intell. Syst. Technol.*, vol.3, no.3, pp.40:1–40:30, 2012.
- [11] J.A. Russell, "A circumplex model of affect," *J. Personality and Social Psychology*, vol.39, no.6, pp.1161–1178, 1980.
- [12] Y. Yamano, R. Cabredo, P. Salvador Inventado, R. Legaspi, K. Moriyama, K.I. Fukui, S. Kurihara, and M. Numao, "Estimating emotions on music based on brainwave analyses," *Proc. 3rd Intl. Workshop on Empathic Computing (IWEC2012)*, pp.115–124, 2012.
- [13] S. Sanei and J. Chambers, *EEG Signal Processing*, Wiley, 2008.
- [14] L.A. Schmidt and L.J. Trainor, "Frontal brain electrical activity EEG distinguishes valence and intensity of musical emotions," *Cognition & Emotion*, vol.15, no.4, pp.487–500, 2001.
- [15] T. Baumgartner, M. Esslen, and L. Jäncke, "From emotion perception to emotion experience: Emotions evoked by pictures and classical music," *Intl. J. Psychophysiology*, vol.60, no.1, pp.34–43, 2006.
- [16] P.J. Lang, M.M. Bradley, and B.N. Cuthbert, "International affective picture system (IAPS): Affective ratings of pictures and instruction manual," Tech. Rep. A-8, The Center for Research in Psychophysiology, University of Florida, Gainesville, FL, 2008.
- [17] G. Chanel, J. Kronegg, D. Grandjean, and T. Pun, "Emotion assessment: Arousal evaluation using EEG's and peripheral physiological signals," *Multimedia Content Representation, Classification and Security*, Lecture Notes in Computer Science, vol.4105, pp.530–537, Springer Berlin Heidelberg, 2006.
- [18] X.-W. Wang, D. Nie, and B.-L. Lu, "Emotional state classification from EEG data using machine learning approach," *Neurocomputing*, vol.129, pp.94–106, 2014.
- [19] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "DEAP: A database for

- emotion analysis using physiological signals," *IEEE Trans. Affective Computing*, vol.3, no.1, pp.18–31, 2012.
- [20] O. AlZoubi, R.A. Calvo, and R.H. Stevens, "Classification of EEG for affect recognition: An adaptive approach," *AI 2009: Advances in Artificial Intelligence, Lecture Notes in Computer Science*, vol.5866, pp.52–61, Springer Berlin Heidelberg, 2009.
- [21] S.A. Hosseini, M.A. Khalilzadeh, M.B. Naghibi-Sistani, and V. Niazmand, "Higher order spectra analysis of EEG signals in emotional stress states," *Proc. 2nd Intl. Conf. on Inf. Tech. and Comp. Sci. (ITCS) 2010*, pp.60–63, 2010.
- [22] P.C. Petrantonis and L.J. Hadjileontiadis, "Emotion recognition from brain signals using hybrid adaptive filtering and higher order crossings analysis," *IEEE Trans. Affective Computing*, vol.1, no.2, pp.81–97, 2010.
- [23] M. Murugappan, N. Ramachandran, and Y. Sazali, "Classification of human emotion from EEG using discrete wavelet transform," *J. Biomedical Science and Engineering*, vol.3, no.4, pp.390–396, 2010.
- [24] D.P.O. Bos, "EEG-based emotion recognition the influence of visual and auditory stimuli," *Capita Selecta*, University of Twente, 2006.
- [25] M.M. Bradley and P.J. Lang, "International affective digitized sounds (IADS): Stimuli, instruction manual and affective ratings," *Tech. Rep. B-2*, The Center for Research in Psychophysiology, University of Florida, Gainesville, FL, 1999.
- [26] R. Khosrowabadi, A. Wahab, K.K. Ang, and M. Baniasad, "Affective computation on EEG correlates of emotion from musical and vocal stimuli," *Proc. Intl. Joint Conf. on Neural Networks (IJCNN) 2009*, pp.1590–1594, 2009.
- [27] Y.-P. Lin, C.-H. Wang, T.-P. Jung, T.-L. Wu, S.-K. Jeng, J.-R. Duann, and J.-H. Chen, "EEG-based emotion recognition in music listening," *IEEE Trans. Biomedical Engineering*, vol.57, no.7, pp.1798–1806, 2010.
- [28] O. Sourina and Y. Liu, "A fractal-based algorithm of emotion recognition from EEG using arousal-valence model," *Proc. Biosignals 2011*, pp.209–214, 2011.
- [29] A. Delorme, T. Mullen, C. Kothe, Z.A. Acar, N. Bigdely-Shamlo, A. Vankov, and S. Makeig, "EEGLAB, SIFT, NIFT, BCILAB, and ERICA: New tools for advanced EEG processing," *Comp. Intell. Neurosci.*, vol.2011, pp.1–12, 2011.
- [30] T.-P. Jung, S. Makeig, C. Humphries, T.-W. Lee, M.J. Mckeown, V. Iragui, and T.J. Sejnowski, "Removing electroencephalographic artifacts by blind source separation," *Psychophysiology*, vol.37, no.2, pp.163–178, 2000.
- [31] Y. Liu and O. Sourina, "EEG databases for emotion recognition," *Proc. Intl. Conf. on Cyberworlds 2013*, pp.302–309, 2013.
- [32] T. Higuchi, "Approach to an irregular time series on the basis of the fractal theory," *Physica D*, vol.31, no.2, pp.277–283, 1988.
- [33] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I.H. Witten, "The weka data mining software: An update," *SIGKDD Explor. Newsl.*, vol.11, no.1, pp.10–18, 2009.
- [34] Y.-P. Lin, Y.-H. Yang, and T.-P. Jung, "Fusion of electroencephalogram dynamics and musical contents for estimating emotional responses in music listening," *Front. Neurosci.*, vol.8, no.94, 2014.
- [35] A. Konar and A. Chakraborty, *Emotion Recognition: A Pattern Analysis Approach*, Wiley, 2015.
- [36] O. Sourina, Y. Liu, and M.K. Nguyen, "Real-time EEG-based emotion recognition for music therapy," *J. Multimodal User Interfaces*, vol.5, no.1–2, pp.27–35, 2012.
- [37] B. Weiss, Z. Clemens, R. Bódizs, and P. Halász, "Comparison of fractal and power spectral EEG features: Effects of topography and sleep stages," *Brain Research Bulletin*, vol.84, no.6, pp.359–375, 2011.
- [38] M. Bachmann, J. Lass, A. Suhhova, and H. Hinrikus, "Spectral asymmetry and higuchi's fractal dimension measures of depression electroencephalogram," *Comp. Math. Methods in Medicine*, vol.2013, pp.299–309, 2013.

Nattapong Thammasan received a B. Computer Eng. degree from Chulalongkorn University in 2012 and an M.Sc. from Osaka University in 2015. He is currently a Ph.D. candidate at the Institute of Scientific and Industrial Research (ISIR), Osaka University. His research interests include artificial intelligence, brain-computer interaction, and affective computing.

Koichi Moriyama received B.Eng., M.Eng., and D.Eng from Tokyo Institute of Technology in 1998, 2000, and 2003, respectively. After working at Tokyo Institute of Technology and Osaka University, he is currently an associate professor at Graduate School of Engineering, Nagoya Institute of Technology. His research interests include artificial intelligence, multi-agent systems, game theory, and cognitive science. He is a member of the Japanese Society for Artificial Intelligence (JSAI).

Ken-ichi Fukui is an Associate Professor in ISIR, Osaka University. He received Master of Arts from Nagoya University in 2003 and Ph.D. in information science from Osaka University in 2010. He was a Specially Appointed Assistant Professor in ISIR, Osaka University from 2005 to 2010, and an Assistant Professor from 2010 to 2015. His research interest includes data mining algorithm and its environmental contribution. He is a member of JSAI, IPSJ, and the Japanese Society for Evolutionary Computation.

Masayuki Numao is a professor in the Department of Architecture for Intelligence, the ISIR, Osaka University. He received a B.Eng. in electrical and electronics engineering in 1982 and his Ph.D. in computer science in 1987 from the Tokyo Institute of Technology. He worked in the Department of Computer Science, Tokyo Institute of Technology from 1987 to 2003 and was a visiting scholar at CSLI, Stanford University from 1989 to 1990. His research interests include artificial intelligence, machine learning, affective computing and empathic computing. He is a member of the Information Processing Society of Japan, the JSAI, the Japanese Cognitive Science Society, the Japan Society for Software Science and Technology, and the American Association for Artificial Intelligence.